Predictive Analytics Using R and Python
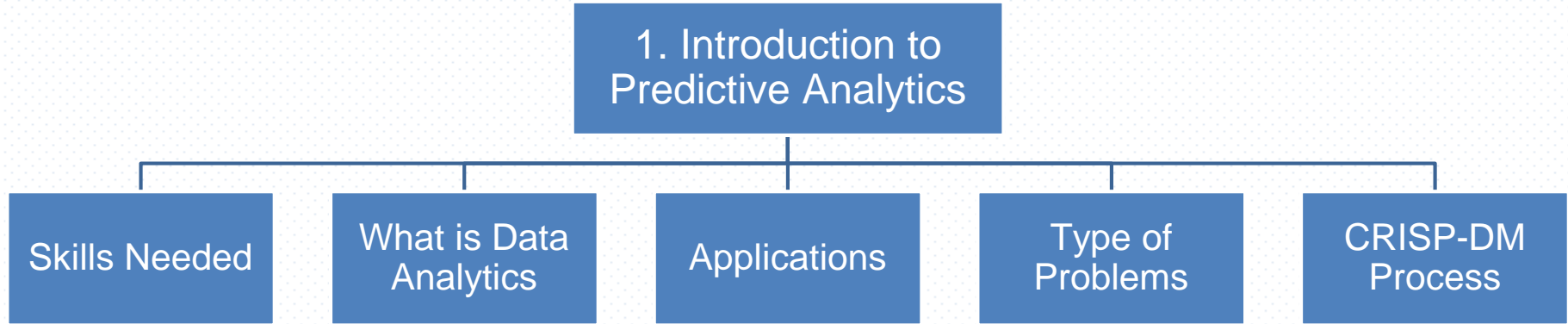
- MUQUAYYAR AHMED
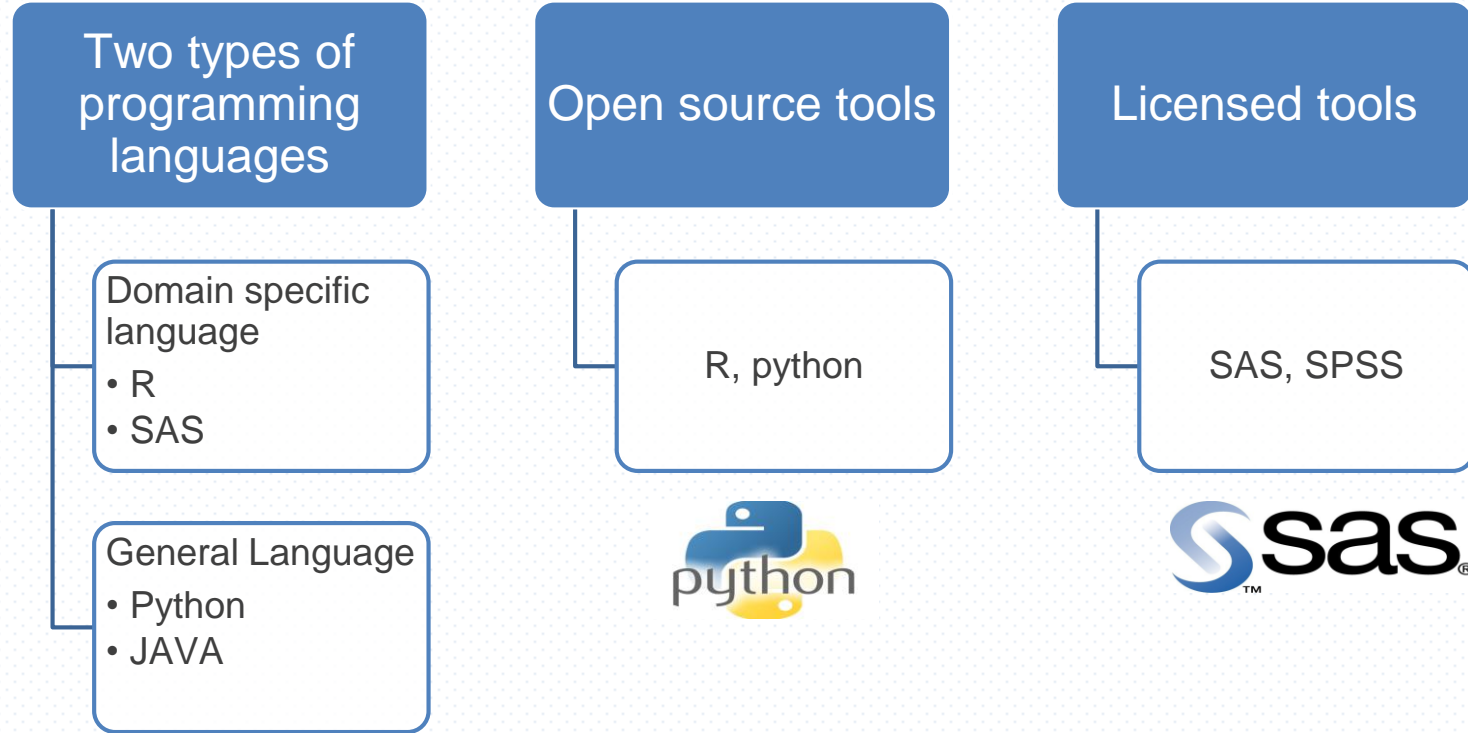**DATA SCIENTIST**

We learnt!!

```
                    ┌─────────────────────┐
                    │ 1. Introduction to  │
                    │ Predictive Analytics│
                    └─────────────────────┘
```

Skills Needed	What is Data Analytics	Applications	Type of Problems	CRISP-DM Process

Today's Agenda

Introduction to R

Two types of programming languages	Open source tools	Licensed tools

Domain specific language
- R
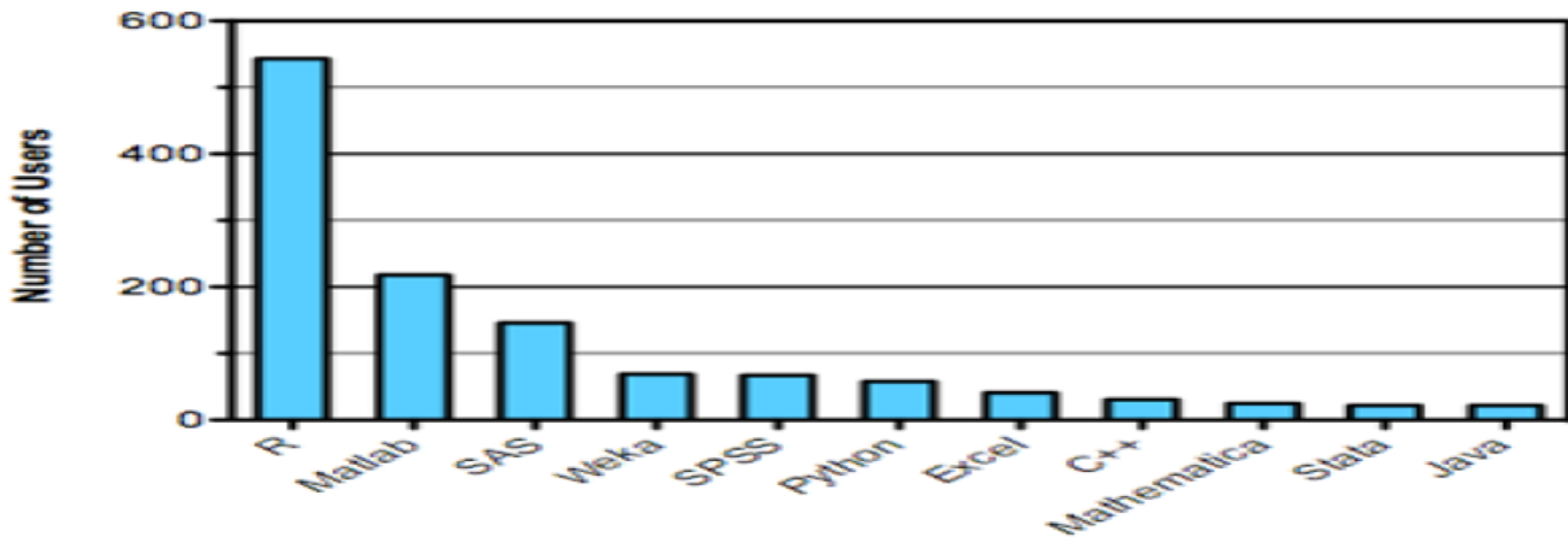- SAS

General Language
- Python
- JAVA

R, python

SAS, SPSS
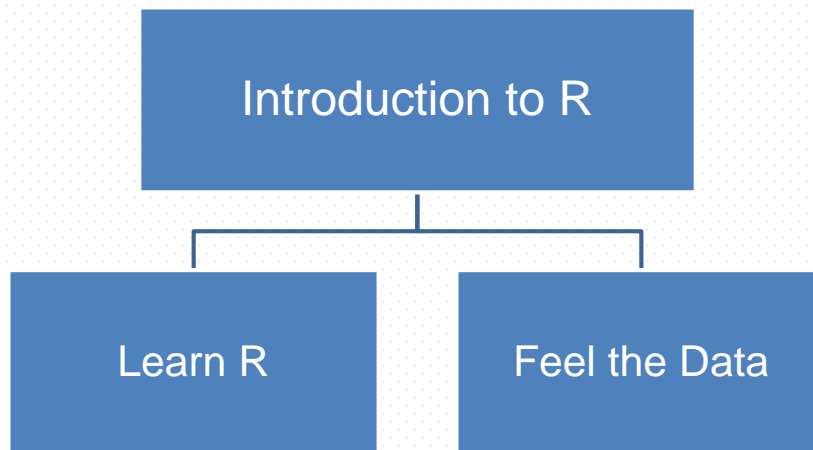
Learn R

Is it worth???

Let's feel real time data and it's jargon

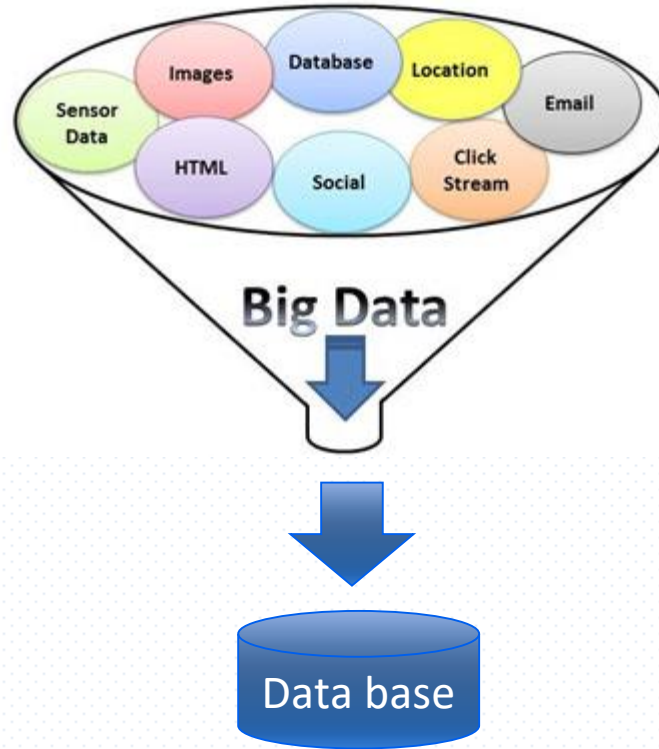Data is limitless, can present anywhere in the universe.

Two types of data
- Primary data
- Secondary data

Database (Storage unit) can be SQL, MySQL, NoSQL, Hadoop clusters.

Data can be in CSV, Excel spread sheets, text files, images, json files etc..

Different sources of data

Sample Data

Two types of variables

- Dependent variable / Target Variable
- Independent variables

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5	3.4	1.5	0.2	Versicolor
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa